



Pinder, M. (2017). On Strawson's critique of explication as a method in philosophy. *Synthese*. <https://doi.org/10.1007/s11229-017-1614-6>

Peer reviewed version

Link to published version (if available):  
[10.1007/s11229-017-1614-6](https://doi.org/10.1007/s11229-017-1614-6)

[Link to publication record in Explore Bristol Research](#)  
PDF-document

This is the author accepted manuscript (AAM). The final published version (version of record) is available online via Springer at <https://link.springer.com/article/10.1007%2Fs11229-017-1614-6> . Please refer to any applicable terms of use of the publisher.

## University of Bristol - Explore Bristol Research

### General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:  
<http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>

# On Strawson's Critique of Explication as a Method in Philosophy

Mark Pinder

University of Bristol

This is a post-peer-review, pre-copyedit version of an article published in *Synthese*. The final authenticated version is available online at: <http://dx.doi.org/10.1007/s11229-017-1614-6>.

**Abstract.** In the course of theorising, it can be appropriate to replace one concept—a folk concept, or one drawn from an earlier stage of theorising—with a more precise counterpart. The best-known account of concept replacement is Rudolf Carnap's 'explication'. P.F. Strawson famously critiqued explication as a method in philosophy. As the critique is standardly construed, it amounts to the objection that explication is 'irrelevant', fails to be 'illuminating', or simply 'changes the subject'. In this paper, I argue that this is an unfair characterisation of Strawson's critique, spelling out the critique in more detail and showing that, fully understood, it is not undermined by extant responses. In light of both the critique and extant responses, I close by making some substantive comments about what explication can, and cannot, be used to do in philosophy.

## 1. Introduction

In the course of theorising, it can be appropriate to replace one concept—a folk concept, or one drawn from an earlier stage of theorising—with a more precise counterpart. For example, on discovering that there are many rocky and icy objects with orbits and sizes quite similar to those of Pluto, including an object, Eris, more massive than Pluto, astronomers deemed the everyday concept of planet to be inadequate. The concept lacked an accepted definition, making it unclear whether or not the discovery of Eris constituted the discovery of a tenth planet. Eris, it seemed, was a borderline case. As such, astronomers explicitly defined a new concept of planet to replace its predecessor—a concept that definitively excluded Eris, as well as Pluto and other similar objects.<sup>1</sup>

---

<sup>1</sup> See International Astronomical Union 2006.

An account of concept replacement, or *explication*, was famously developed by Rudolf Carnap.<sup>2</sup> He defended the use of explication in science, logic, mathematics and philosophy. In by far the most influential critique of Carnap's account of explication, P.F. Strawson criticises the claim that explication can help to solve *philosophical* problems.<sup>3</sup> As the critique is standardly construed, two passages are given particular weight. Firstly:

[...] to offer formal explanations of key terms of scientific theories to one who seeks philosophical illumination of essential concepts of non-scientific discourse, is to do something utterly irrelevant—is a sheer misunderstanding, like offering a text-book on physiology to someone who says (with a sigh) that he wished he understood the workings of the human heart. (Strawson 1963: 505)

Secondly:

[...] typical philosophical problems about the concepts used in non-scientific discourse cannot be solved by laying down the rules of use of exact and fruitful concepts in science. To do this last is not to solve the typical philosophical problem, but to change the subject. (p. 506)

The underlying objection is typically glossed as being that, when it comes to philosophy, explication is 'irrelevant', fails to be 'illuminating', or simply 'changes the subject'.<sup>4</sup>

This is not, however, a fair characterisation of Strawson's critique. The critique is more subtle than is generally acknowledged and, consequently, is not undermined by extant responses. More importantly, once the critique has been spelt out, we are better placed to delimit the role that explication can play in philosophical methodology.

My aim in this paper is twofold. First, I aim to correct our understanding of Strawson's critique of explication as a method in philosophy. Second, in light of both the critique and extant

---

<sup>2</sup> See e.g. Carnap 1947; 1950.

<sup>3</sup> See Strawson 1963.

<sup>4</sup> See e.g. Brun 2016: 1219; Dutilh Novaes and Reck 2017: 202; Justus 2012: 170–171; Loomis and Juhl 2006; Maher 2007: 332; Olsson 2015: 69; Schupbach 2015: 12–13.

responses, I make some substantive comments about what explication can, and cannot, be used to do in philosophy.

## 2. Explication

Explication is the replacement of one or more concepts, the *explicandum* or *explicanda*, with more precise counterparts, the *explicatum* or *explicata*. Explicanda can be folk concepts or concepts from an earlier stage of theorising; and explicata can (but need not) be expressed using the same words as corresponding explicanda, as in the case of PLANET.<sup>5</sup>

An explicatum is to satisfy four requirements “to a sufficient degree” (1950: 7). Firstly, the explicatum is to be *similar* to the explicandum “in such a way that, in most cases in which the explicandum has so far been used, the explicatum can be used” (*ibid.*).<sup>6</sup> Secondly, the explicatum is to be *precise*, or at least *more* precise than the explicandum, in virtue of explicit rules of use (or definition) given in terms of a “well-connected system” of concepts (*ibid.*). Thirdly, the explicatum is to be *fruitful*, in the sense that it facilitates the “formulation of many universal statements (empirical laws in the case of a nonlogical concept, logical theorems in the case of a logical concept)” (*ibid.*). Fourthly, the explicatum is to be *simple*.

Here is an example given by Carnap (pp. 12–15). It involves the replacement of the prescientific concept WARMER, understood to depend solely on our sensations, with the explicatum TEMPERATURE, understood as a quantitative concept. The four requirements are satisfied as follows. First, similarity: in most cases in which  $x$  is (according to our sensations) warmer than  $y$ , the temperature of  $x$  is greater than the temperature of  $y$ . Second, precision: rules of use for

---

<sup>5</sup> Here and below I use small capitals to denote concepts.

<sup>6</sup> This is *not* best interpreted in terms of the extensions of explicanda and explicata. Rather, it is better to construe similarity as requiring an explicatum to be suitable for performing at least the theoretical work performed by the corresponding explicandum (and perhaps more). For example, as Brun notes (2016: 1221), one might explicate the concept PROPOSITION with the concept SET OF POSSIBLE WORLDS, even if one thinks that the extension of PROPOSITION does not contain any sets of possible worlds. Here, the concepts are similar in the sense that, relative to the specific context in which the explicatum is used, the explicatum “performs the explicandum’s function” (Brun 2016: 1222).

TEMPERATURE can be precisely defined with reference to thermometers. Third, fruitfulness: TEMPERATURE features in (for example) the ideal gas law. And, fourth, simplicity: both the rules for the use of TEMPERATURE, and the laws in which it features, are simple. In light of such considerations, Carnap takes TEMPERATURE to be “the [explicatum of WARMER] important for science” (p. 14).

The explicatum replaces the explicandum in relevant theoretical contexts. Thus, in theoretical contexts in which physicists (*qua* physicists) would have used the explicandum WARMER, they ought now to instead use the explicatum TEMPERATURE. This will not, in general, be a simple instance-for-instance replacement—trivially so in the present case given that WARMER, but not TEMPERATURE, is a comparative concept. An explicatum replaces an explicandum in a theoretical context in the sense that, in that context, the explicatum will do at least the theoretical work that was done by the explicandum (and possibly more).<sup>7</sup>

Some comments are in order. First, explication is sometimes characterised as introducing *formal* systems of concepts.<sup>8</sup> For example, Strawson writes that

[the] method is to construct a formal system, which uses, generally, the ordinary apparatus of modern logic and in which the concepts forming the subject-matter of the system are introduced by means of axioms and definitions. (1963: 503)

However, as has been made clear by Brun (2016: 1217f) and Dutilh Novaes and Reck (2017: 200f), this should not be thought of as a requirement. Rather, as Carnap writes in *Meaning and Necessity*, explication involves “replacing [a vague or not quite exact concept] by a newly constructed, *more* exact concept” (1947: 8, my emphasis). One advantage of understanding explication in this way is that it better reflects actual scientific and philosophical practice, which—as illustrated by the opening example of PLANET—often involves the construction of concepts that are informal or, at least, not fully formalised.

---

<sup>7</sup> See e.g. Brun 2016: 1218.

<sup>8</sup> Suggestive comments along these lines can be found in, for example, Maher 2007, Schupbach 2015, and Strawson 1963.

Second, Carnap is typically held to have prioritised the fruitfulness requirement over similarity, precision and simplicity. Thus, for example, Schupbach writes that “Carnap plays favorites with regards to his desiderata, prioritizing fruitfulness over similarity” (2015: 7) and Dutilh Novaes and Reck write that “fruitfulness is ultimately the most significant requirement for an explication overall” (2017: 202). The thought, as I understand it, is twofold. First, a concept requires a sufficient degree of fruitfulness, similarity, precision and simplicity to count as an explicatum. Second, given multiple candidate explicata—that is, concepts with a sufficient degree of fruitfulness, similarity, precision and simplicity—fruitfulness is prioritised as the most important factor in determining which of those candidates to choose. Importantly, though, we ought not to build the prioritisation of fruitfulness into the method of explication *per se*. An explication is performed so long as the explicandum is replaced with a candidate explicatum, whichever factors one subsequently prioritises in determining which candidate explicatum to choose.

Third, as I have characterised explication, it involves not merely the introduction of explicata but also the *replacement* of explicanda. Thus, if a theorist constructs an explicatum but goes on to use both it and the explicandum in relevant theoretical contexts, then she has not fully explicated the explicandum. If a physicist (*qua* physicist) continues to find it convenient to use WARMER when (say) theorising about pressure and volume, then she has not fully explicated that concept. I will say that the mere construction and introduction of explicata is a *partial explication*; if the explicanda are also replaced, then it is a *full explication*. For terminological ease, I take both partial and full explication to be types of explication.

To avoid misunderstanding, let me clarify two points about the partial/full distinction. First, both types of explication are compatible with explicanda being used *outside* of the relevant theoretical contexts. The physicist who fully or partially explicates WARMER may nonetheless use that concept when (say) she compares the potatoes and carrots over the dinner table. Second, the distinction does not represent a meta-theoretical choice: there are no ‘full-explicationists’ or ‘partial-explicationists’. Rather, when there is specific reason to eliminate the explicanda—perhaps because new discoveries have shown the explicanda not to be fit for their intended purposes, as in the case of PLANET—the

theorist will *fully* explicate the explicanda; lacking any such reason, she may *either* fully or partially explicate the explicanda.

### 3. Strawson's Critique

Strawson's critique is often characterised—or caricatured—as an objection to explication on the simple grounds that explication is 'irrelevant' and 'changes the subject'. However, the critique is more subtle and deserving of careful examination than this characterisation suggests. In this section I develop the critique, leaving critical discussion for subsequent sections.

#### 3.1 Preliminaries

Strawson's stated aim is to discuss the "comparative merits of two methods of philosophical clarification" (1963: 503). The first method is explication (although Strawson calls it 'rational reconstruction'); the second method consists of "the attempt to describe the complex patterns of logical behaviour which the concepts of daily life exhibit" (*ibid.*), and falls under the umbrella of 'ordinary language philosophy'. I focus exclusively on Strawson's consideration of the merits of explication.

Although Strawson is not explicit, he is naturally interpreted as taking *philosophical clarification* (or *philosophical illumination*) to consist of the clarification of concepts for the purposes of solving philosophical problems and answering philosophical questions. The central type of philosophical problem that he highlights is "dealing with paradox and perplexity" (p. 515):<sup>9</sup>

For it often happens that someone reflecting on a certain set of concepts finds himself driven to adopt views which seem [...] paradoxical or unacceptably strange. (*ibid.*)

---

<sup>9</sup> Strawson distinguishes two additional, interrelated types of philosophical problem. The first is "the attempt to explain [...] why it is that we have such concepts and types of discourse as we do" (pp. 515–516). This is not a historical inquiry, but rather an inquiry into why, given our natures, it is natural for us to have the particular conceptual framework that we do. The second is the "examination of current concepts and types of discourse [...] with no particular therapeutic purpose, but for its own sake" (p. 517). Herein, I focus on paradoxes and perplexity.

The idea is familiar. Consider, for example, the following argument: *some dreams are indistinguishable from sensory experience, so my experience is compatible with my being in such a dream, so I don't know that I'm not dreaming and so, finally, I don't know that I am currently (say) sitting in my armchair.* When reflecting on the concepts involved, especially KNOWLEDGE, the argument seems to be sound but the conclusion (or some appropriate variant thereof) seems to be unacceptable: surely I *do* know that I am currently sitting in my armchair (or that I am currently working at my desk, or whatever). Let us call such an argument—an argument such that, when reflecting on the concepts involved, the argument seems sound and the conclusion seems unacceptable—a *conceptual imbalance*. Let us call the specific example just given the *sceptical imbalance*.

To solve a philosophical problem,

the critical philosopher must not only restore the conceptual balance which has somehow been upset; he must also diagnose the particular sources of the loss of balance, show just how it has been upset. (*ibid.*)

Thus, in response to the sceptical imbalance, it would not be enough (say) to simply insist that I do know that I am not dreaming. One would also have to explain why appearances are to the contrary—why it *seems* that, given the compatibility of my experience and a vivid dream, I do *not* have such knowledge. More generally, for Strawson, one must do two things to respond to a conceptual imbalance: first, one must explain why a given step in the argument is in fact unsound, or why the conclusion is in fact acceptable; second, one must explain why appearances are to the contrary. I will say that to respond to a conceptual imbalance in this way is to *resolve* it. So, according to Strawson, one *solves* a philosophical problem by *resolving* the corresponding conceptual imbalance.

It is worth noting that this account of philosophical problem presupposes that we have not adopted an externalist semantics for concepts that are the target of explication.<sup>10</sup> For example, suppose in the essentialist tradition of Putnam and Kripke that the semantic content of an explicandum is the

---

<sup>10</sup> Thanks to an anonymous referee for helping me to see this.



essence of whatever bears the appropriate causal-historical relation to uses of that explicandum.<sup>11</sup>

Then the kind of ‘paradoxes and perplexities’ Strawson has in mind do not arise from reflection on the concept involved, but rather from confusion about the extra-conceptual world: the problem of scepticism does not require clarification of KNOWLEDGE, but clarification of knowledge.<sup>12</sup> For this reason, let us put semantic externalism aside for the present. I return to the issue in §5.3.

Strawson considers two ways of understanding explication, examining for each how it might be used for philosophical clarification in the above sense. According to the first (pp. 504–506), explication is “the introduction, for scientific purposes, of scientifically exact and fruitful concepts” (p. 504). According to the second (pp. 510–514), explication is the construction of precise systems of concepts such that “at least some of the key concepts in the system are, in important respects, very close to the ordinary concepts which are to be clarified” (p. 511).

It is helpful to understand this in light of a distinction drawn by Jonah Schupbach (2017: 676–680). Schupbach considers how two of the requirements for explicata—similarity and fruitfulness—are to be weighted. The answer, he suggests, depends on one’s aims. If, following Schupbach’s interpretation of Carnap, one’s aim is “the improvement of our concepts in the service of the empirical or logicomathematical sciences” (p. 680), then one will prioritise fruitfulness over similarity: “one’s explicatum is satisfactory to the extent that it is *as similar to the explicandum as fruitfulness allows*” (p. 677). In contrast if, following Schupbach’s interpretation of Kemeny and Oppenheim (1952), one aims to “shed new light on unclear concepts by introducing corresponding concepts more readily analyzable” (Schupbach 2017: 680), then one should prioritise similarity over fruitfulness. From this perspective, “the explicatum is satisfactory to the extent that it is *as fruitful as the more fundamental desideratum similarity allows*” (p. 679). Thus, according to Schupbach, we can distinguish between two types of explication. The first, *Carnapian explication*, seeks to build precise explicata that are as fruitful in the service of the (empirical or logicomathematical) sciences as possible;<sup>13</sup> the second,

---

<sup>11</sup> See Kripke 1980; Putnam 1975.

<sup>12</sup> Cf. Williamson 2007.

<sup>13</sup> I put aside the exegetical question of whether Carnapian explication (so defined) is Carnap’s conception of explication.

*Oppenheimian explication*, seeks to build precise explicata that are, in relevant respects, as similar to the explicanda as possible.

This distinction maps onto Strawson's. Carnapian explication is, broadly speaking, 'the introduction, for scientific purposes, of scientifically exact and fruitful concepts'; and Oppenheimian explication is such that 'at least some of the key concepts in the constructed system will be very close to the ordinary concepts which are to be clarified'. Thus, with the distinction so construed, we can see Strawson's critique of explication as having two parts: the first involves an examination of the use of *Carnapian* explication in solving philosophical problems; and the second involves an examination of the use of *Oppenheimian* explication in solving philosophical problems.

### **3.2 Strawson on Carnapian Explication**

Strawson's examination of Carnapian explication includes the rhetorical quotations in the opening section of this paper. He suggests that it is "prima facie evident" that Carnapian explication is "utterly irrelevant" to philosophical clarification (1963: 505), serving only to "change the subject" (p. 506). It is worth spelling out Strawson's position in more detail.

Consider two passages. First:

The scientific uses of language, whether formal or empirical, are extremely highly specialized uses. Language has many other employments. We use it in pleading in the law courts; in appraising people's characters and actions; [...]; in describing how things look and sound and feel like; [...]; and so on. It is quite certain that such ways of using language as these may give rise to philosophical problems; that the concepts employed in these activities may generate perplexity, may call for philosophical clarification. (p. 505)

There are two things to note here. First, Strawson takes the everyday use of language to be broader than the scientific use of language. Second, he takes the everyday use of language and concepts to give rise to philosophical problems. The second passage:

[...] the concepts used in non-scientific kinds of discourse could not literally be *replaced* by scientific concepts serving just the same purposes; [...] the language of science could not in

this way *supplant* the language of the drawing-room, the kitchen, the law courts and the novel.  
[...] The kinds of concept we employ are not independent of the kinds of purpose for which  
we employ them [...]. (pp. 505–506)

Part of Strawson's thought here is that, given that the use of language is narrow in scientific contexts and broad in everyday contexts, the concepts introduced for use in science must serve a different (narrower) purpose than those deployed in everyday contexts—which is why the former cannot straightforwardly supplant the latter.

The upshot, not quite made explicit, is this: the concepts used in everyday contexts give rise to a particular set of philosophical problems; those introduced for use in scientific contexts, as a result of their having narrower purposes, do *not* give rise to that particular set of philosophical problems.

By way of example, Strawson alludes to Carnap's example of WARMER and TEMPERATURE:

Sensory concepts in general have been a rich source of philosophical perplexity. [...] Does it follow from the fact that the same object can feel warm to one man and cold to another that the object really is neither cold or warm nor cool nor has any such property? [Such] questions can be answered, or the facts and difficulties that lead to our asking them can be made plain; but not by means of formal exercises in the scientific use of the related concepts of temperature, wavelength, frequency. (p. 506)

Strawson states, here, that a philosophical problem arises in connection with the question posed in the quotation. He may have in mind something like the following argument: *this plate feels warm to Jessica but feels cold to Luke, but warmth and coldness are incompatible properties and there's no reason to say that just one of Jessica and Luke is wrong, so the plate cannot genuinely be warm or cold nor have any other such property*. The argument is plausibly such that, when reflecting in the concepts involved, the argument seems sound while the conclusion seems unacceptable. Assuming so, this is a conceptual imbalance; call it the *warmth imbalance*.

Strawson's claim in the passage is that we do not resolve the warmth imbalance by replacing WARM, COLD and related concepts with the more scientifically fruitful concept TEMPERATURE. Building on the above comments, the principal reason is that scientifically fruitful concepts tend not

to give rise to the kind of conceptual imbalance at hand. First, WARM, COLD, etc., are used for a greater variety of purposes than TEMPERATURE: for example, the former are perhaps used *both* to track an objective quality of objects and to reflect our sensations thereof, whereas the latter is perhaps used *merely* to track an objective quality of objects.<sup>14</sup> This narrowing of purpose prevents the warmth imbalance from arising. After replacing WARM, COLD, etc., with TEMPERATURE, arguments designed to generate the warmth imbalance are obviously unsound. For example: *this plate feels to Jessica to have a high temperature, but feels to Luke to have a low temperature, but having a high temperature and having a low temperature are incompatible properties and there's no reason to say that just one of Jessica and Luke is wrong, so the plate cannot genuinely have a temperature.* The error in the argument is clear: TEMPERATURE (defined in terms of thermometers) need not reflect one's sensations, so either Jessica or Luke is straightforwardly mistaken. So, once WARM, COLD, etc., have been replaced by TEMPERATURE, the warmth imbalance no longer arises: there is no longer a seemingly sound argument for a seemingly unacceptable conclusion.

But none of this serves to resolve the warmth imbalance. Firstly, it does not automatically follow from the fact that TEMPERATURE need not reflect one's sensations, that WARM, COLD, etc., also need not reflect one's sensations: we cannot simply infer that, because Jessica or Luke is straightforwardly mistaken about the plate's temperature, either Jessica or Luke must likewise be straightforwardly mistaken about the plate being warm/cold. Moreover, even if such an inference *could* be justified, highlighting a step in the argument that generates the warmth imbalance as fallacious would not tell us why that step seems to be sound. Replacing WARM, COLD, etc., with TEMPERATURE would not serve to explain why there is something intuitively amiss—assuming that

---

<sup>14</sup> An anonymous referee suggests that there might in fact be *two* concepts denoted by 'WARM': one that tracks an objective property, and one that tracks a subjective property. If this is right, then, on either reading, the argument used to set up the warmth imbalance is obviously unsound: on the objective reading, there *is* reason to say that just one of Jessica and Luke is wrong; on the subjective reading, warmth and coldness are not incompatible. This insight would potentially allow us to resolve the warmth imbalance (without appealing to Carnapian explication), by explaining why the two distinct concepts are sometimes conflated. For expository purposes, however, I put this possibility aside; I assume that 'WARM' denotes a single concept.

there *is* something intuitively amiss—with the thought that one can be straightforwardly mistaken about whether a plate is warm or cold. Rather than resolving the warmth imbalance, Carnapian explication replaces the explicanda with a set of concepts for which the conceptual imbalance does not arise. And as Justus writes, “[d]esertion does not constitute illumination” (2012: 171).

Let me summarise Strawson’s argument here. A philosophical problem is solved by resolving the corresponding conceptual imbalance. Conceptual imbalances arise due to some aspect of the interaction between the different purposes for which a concept is used. But, when one performs a Carnapian explication, one replaces that concept with a counterpart that is explicitly designed to be used for much narrower purposes. As such, if one attempts to reconstruct the conceptual imbalance using relevant explicata, one will see that the conceptual imbalance no longer arises: either the argument does not seem to be sound, or the conclusion does not seem to be unacceptable. However, performing this Carnapian explication—and thereby blocking the conceptual imbalance—does not *resolve* the conceptual imbalance. The Carnapian explication has not told us why the unsound step in the original argument seems to be sound, or else how or why the acceptable conclusion seems to be unacceptable—why, as it were, reflection on the concepts involved can lead us astray. Thus, Strawson concludes, Carnapian explication fails to solve the philosophical problem.

Let me make three points of clarification. First, it is essential here that we are talking about *Carnapian* explication rather than *Oppenheimian* explication. For Strawson, constructing a concept for scientific fruitfulness will narrow down its purpose, thereby removing the conceptual imbalances that are central to his conception of philosophical problems, and it is this removal that underpins his criticism. There is no reason to think that a parallel claim could be made about an explicatum designed to maximise *similarity* to the explicandum.

Second, for the objection to have any force, we must have in mind Carnapian *full explication* rather than Carnapian *partial explication*. It is the *replacement* of WARM, COLD, etc., with TEMPERATURE, not the mere introduction of the latter, that prevents us from formulating the philosophical problem. The introduction of TEMPERATURE does not remove the conceptual imbalance, nor does it hinder any attempt we might make to resolve it. It is the elimination of WARM, COLD, etc., that causes the difficulty.

Third, there is no essential distinction here between ‘everyday’ and ‘scientific’ concepts, contexts, language, or anything else.<sup>15</sup> What is important for the objection is just that Carnapian explication, in designing concepts for fruitfulness in the service of science, will yield explicata that do not give rise to the particular conceptual imbalances to which the explicanda are known to give rise. This claim does not require a problematic, sharp distinction between the everyday and the scientific.

### 3.3 Strawson on Oppenheimian Explication

Oppenheimian explication, recall, seeks to maximise how similar the explicatum is to the explicandum in some relevant respects. According to Strawson, Oppenheimian explications are to be accompanied with “extra-systematic remarks” in order to “explain the meaning of the linguistic expressions for the constructed concepts in terms which do not belong to the theory” (1963: 512). This is particularly important when we have philosophical problems in mind:

[...] if the clear mode of functioning of the constructed concepts is to cast light on problems and difficulties rooted in the unclear mode of functioning of the unconstructed concepts, then precisely the ways in which the constructed concepts are connected with and depart from the unconstructed concepts must be plainly shown. (p. 513)

And this, according to Strawson, can only be achieved by “accurately describing the modes of functioning of the unconstructed concept” (*ibid.*). Moreover,

[...] in so far as the purpose of a constructed system is philosophical clarification, the extra-systematic remarks, so far from being [...] comparatively unimportant trimmings, are just what give life and meaning to the whole enterprise. (*ibid.*)

That is, if one uses Oppenheimian explication to solve a philosophical problem, then it is the extra-systematic remarks, not the explication *per se*, that do the work.

For example, suppose that a theorist aims to resolve the warmth imbalance. She might proceed by offering an Oppenheimian explication, constructing a system of concepts, WARM', COLD',

---

<sup>15</sup> Cf. Carnap 1963.

etc., that are in relevant respects very similar to the explicanda. Perhaps, say, an object  $x$  falls under WARM' or COLD' at  $t$  if the typical healthy human at room temperature would, upon contact with  $x$  at  $t$ , have one or another particular kind of sensation.<sup>16</sup> The theorist might then use the explicata, in some way or another, to resolve the warmth imbalance. First, she might point out that, if these explicata are used to reconstruct the argument that generates the warmth imbalance, the reconstructed argument is unsound: there *is* reason to say that just one of Jessica and Luke is wrong about the plate being warm/cold. Second, she might argue that it may nonetheless *seem* that neither Jessica nor Luke are wrong, because both Jessica and Luke may have an equal *prima facie* claim to representing the typical healthy human. Thus, the theorist may conclude that the warmth imbalance arises because, say, we are not sufficiently sensitive to the fact that different healthy humans can have different sensations in parallel circumstances.

The details, here, are not important. According to Strawson, the pivotal step—the step that ‘gives life and meaning’ to any such resolution—is the *provision of an explanation* of exactly how the explicata are similar to the explicanda. The idea is this. First, the theorist’s resolution of the warmth imbalance will turn on her ability to explain exactly how WARM' and COLD' are similar to WARM and COLD. And this explanation, Strawson claims, must make close reference to the behaviour of the explicanda. But then, so goes the thought, it is the theorist’s understanding of this behaviour that enables her to resolve the warmth imbalance: she could resolve the warmth imbalance simply by reflecting on her explanation of how WARM and COLD are tied to judgements about the typical healthy human’s sensations. For Strawson, the additional step of constructing and introducing the explicata is unnecessary. It is the *explanation*—rather than the *explication*—that does the work in philosophically clarifying the explicanda; the explication itself “must remain ancillary” (p. 517).

There are two points to make explicit. First, Strawson is *not* arguing that Oppenheimian explication is irrelevant to solving philosophical problems, nor is he claiming that it changes the

---

<sup>16</sup> As things stand, WARM' and COLD' are probably not sufficiently precise to count as candidate explicata. The theorist of the example would need to provide more precise definitions—perhaps defining the sensations demonstratively, providing a more concrete account of ‘the typical healthy human’, and so on.

subject. Strawson accepts that Oppenheimian explication may help solve philosophical problems; he simply downplays its importance. Second, in contrast to before, it is inessential here that we are talking about Oppenheimian *full explication* rather than Oppenheimian *partial explication*. Strawson's critique of Oppenheimian explication does not rely on the replacement of explicanda, but on the comparative importance of the extra-systematic remarks.<sup>17</sup>

### 3.4 Summary

Strawson critically examines two ways in which explication might be used as a method of philosophical clarification. First, if one seeks to use Carnapian full explication to solve a philosophical problem, thereby replacing the concepts used to formulate the philosophical problem with concepts designed to be fruitful for science, then one will change the subject. The reason is that concepts designed to be fruitful for science have narrower purposes than the concepts they are to replace, and thus do not give rise to the philosophical problems at hand. Second, if one seeks to use Oppenheimian explication to solve a philosophical problem, thereby constructing concepts that are maximally similar to the concepts used to formulate the philosophical problem, then it will not be the explication *per se* that does the work but the extra-systematic remarks relating the explicata and explicanda. So when it comes to solving philosophical problems, Strawson concludes that Carnapian full explication changes the subject and Oppenheimian explication is ancillary.

## 4. Extant Responses

There are three prominent lines of response to Strawson's critique. With the critique spelt out in detail, we can see that none of the objections fully undermine that critique. Nonetheless, each response bears upon the critique in an important and instructive way.

---

<sup>17</sup> It is also inessential that we are talking about *Oppenheimian* explication rather than *Carnapian* explication. Strawson could accept that, if Carnapian explication *were* relevant to a philosophical problem, it would nonetheless be the explanation of *how* it is relevant that would 'give life and meaning' to any subsequent solution.



#### 4.1 Pocketknives and Microtomes

In response to Strawson, Carnap states that he is “firmly convinced” that “an explication given, not in the ordinary language, but in a scientific, technically constructed language [can] be useful for the solution of a philosophical problem” (1963: 934), something he takes Strawson to explicitly deny. He gives the following oft-quoted analogy:<sup>18</sup>

A natural language is like a crude, primitive pocketknife, very useful for a hundred different purposes. But for certain specific purposes, special tools are more efficient, e.g., chisels, cutting-machines, and finally the microtome. If we find that the pocket knife is too crude for a given purpose and creates defective products, we shall try to discover the cause of the failure, and then either use the knife more skilfully, or replace it for this special purpose by a more suitable tool, or even invent a new one. [Strawson’s] thesis is like saying that by using a special tool we evade the problem of the correct use of the cruder tool. But would anyone criticize the bacteriologist for using a microtome, and assert that he is evading the problem of correctly using a pocketknife? (pp. 938–939)

There are two brief points to make in response to Carnap.

First, *pace* Carnap, Strawson does not deny that explication *per se* can be useful for solving philosophical problems—he accepts, for example, that Oppenheimian explication can play an ancillary role. Rather, Strawson denies that *replacing* concepts of interest with those designed to be *fruitful for science* can help to solve philosophical problems.

Second, as observed by Patrick Maher, Carnap’s analogy is misplaced:

nobody would criticize the bacteriologist, but that is because the bacteriologist’s problem was not about the pocketknife. However, the relevant analogy for “one who seeks philosophical illumination of essential concepts of non-scientific discourse” is someone who seeks knowledge of proper use of the pocketknife [...]. (2007: 333)

---

<sup>18</sup> See, e.g., Loomis and Juhl 2006; Maher 2007: 323–323; Olsson 2015: 69; Schupbach 2015: 13.

Thus, Carnap's analogy does not tell against Strawson's critique: firstly, one would not gain knowledge of the proper use of a pocketknife by replacing it with a microtome; and, secondly, if somehow one were to gain such knowledge via a microtome, then detailed antecedent knowledge of how pocketknife-use is related to microtome-use would be required.<sup>19</sup>

Nonetheless, Carnap's response does bear on Strawson's critique. Carnap's analogy trades on the thought that, when one is not specifically interested in the explicandum, explication appears to be unproblematic. For example, the astronomers' explication of PLANET is unproblematic in part because they sought to provide a principled taxonomy of celestial bodies, rather than to better understand the source of puzzlement in our everyday conception of planethood. A parallel point applies in some philosophical cases: explication appears to be unproblematic for cases in which one is *not* seeking to philosophically clarify the explicandum.<sup>20</sup> For example, if one seeks to philosophically clarify KNOWLEDGE in light of the sceptical imbalance, then it is a viable strategy to explicate (say) SENSORY EXPERIENCE in setting up the problem. Or, alternatively, if one seeks to provide a principled taxonomy of epistemic states (rather than to resolve the sceptical imbalance), then it is a viable strategy to explicate KNOWLEDGE. These points do not serve to undermine Strawson's critique of what he calls "philosophical problems"; rather, they show that Strawson's understanding of philosophical problems is not exhaustive and that his critique is unlikely to fully generalise. I return to this line of thought in §5.

## 4.2 Arguments by Analogy

A number of authors have responded to Strawson's critique along the following lines.<sup>21</sup> Suppose that we are interested in whether concept *C* has property *F*—i.e. in whether  $F(C)$ —but have been unable to determine an answer directly. Then, we might explicate *C*, resulting in explicatum *C'*. Given that *C'* is

---

<sup>19</sup> As an anonymous referee notes, it is essential to this response that we have put aside semantic externalism—and that we are interested in concepts (the pocketknife) rather than the extra-conceptual world (bacteria). See the comment in §3.1, and see §5.3 for further discussion. Cf. Williamson 2007.

<sup>20</sup> A similar point is made by Brun (2016: 1219) and Schupbach (2017: 685).

<sup>21</sup> See: Brun 2016: 1219; Maher 2007: 333–334; Olsson 2015: 69–70; Schupbach 2017: 691.

more precise than  $C$ , we might be able to establish (say) that  $F(C')$ . Then, pointing to the similarity between  $C$  and  $C'$ , we might argue that, by analogy,  $F(C)$ . Call an argument of this form an *argument by analogy*.

Patrick Maher develops the idea most explicitly (2007: 333–334). He begins by supposing that we seek to determine whether a sentence  $S$  of ordinary language is true: we proceed by replacing the imprecise concepts in  $S$  with explicata, thereby forming a corresponding sentence  $S'$ , and then we determine the truth value of  $S'$ . Maher argues that this can help to determine the truth value of  $S$ :

(1) The attempt to formulate  $S'$  often shows that the original sentence  $S$  was ambiguous or incomplete and needs to be stated more carefully. (2) If the explicata appearing in  $S'$  are known to correspond well to their explicanda in other cases, that is a reason to think that they will correspond well in this case too, and hence to think that the truth value of  $S$  will be the same as that of  $S'$ . (3) We can translate the proof or disproof of  $S'$  into a parallel argument about the corresponding explicanda and see if this seems to be sound; if so, we obtain a direct argument for or against  $S$ . (p. 334)

Maher concludes that “explication can provide insights and lines of argument that we may not discover if we reason only in terms of the vague explicanda” (*ibid.*).

Maher gives an example. Consider the sentence: “A law of the form ‘All  $F$  are  $G$ ’ is confirmed by evidence that something is both  $F$  and  $G$ ”. Second, suppose that we explicate CONFIRMATION, replacing it with a more precise concept CONFIRMATION'. Then, (1), as available formal notions of confirmation are sensitive to background evidence, we will discover that our original sentence is ambiguous. Here is Maher’s disambiguation:

$N$ :     A law of the form ‘All  $F$  are  $G$ ’ is confirmed by evidence that something is both  $F$  and  $G$ ,  
given no background evidence.

We then construct an analogue of  $N$  by replacing “confirmed” with “confirmed'”. Call the analogue  $N'$ . Regarding (2), Maher (assuming his preferred explication of CONFIRMATION) cites a proof that  $N'$

is false (Maher 2004). He claims that, as CONFIRMATION corresponds well with CONFIRMATION' in most cases, we thus have reason to suppose that *N* is false. And (3), Maher claims that, if we translate his proof into “qualitative explicandum terms” (2007: 334), we obtain a good direct argument for the falsity of *N*.

Arguments by analogy look to be a plausible way in which one might use explication to establish results about explicanda. As such, they present an effective response to the simple charge that, when it comes to philosophy, explication is irrelevant. However, as we have seen, Strawson's critique is not to be identified with that simple charge. So let us examine the extent to which arguments from analogy bear on Strawson's critique.

Suppose we have Strawson's critique of Carnapian explication in mind. Then, according to Strawson, it is the *replacement* of explicanda that undermines any attempt to solve philosophical problems. However, arguments by analogy presuppose that replacement has *not* taken place: they presuppose a *partial* explication. To see this, recall that arguments by analogy seek to establish  $F(C)$  by analogy to the independently established  $F(C')$ , where *C* is the explicandum and *C'* is the explicatum. Such an argument must presuppose a theoretical context in which *C'* was designed to do the same kind of theoretical work as *C*—only then is there a relevant analogy between the concepts. But the very fact that *C* is being used in that context demonstrates that, in that context, *C* has not been *replaced* by *C'*: only a *partial* explication has been performed. Thus, arguments by analogy make use of partial, as opposed to full, explications. Strawson's critique, however, is silent about Carnapian *partial* explications. As such, with Carnapian explications in mind, arguments by analogy are straightforwardly compatible with Strawson's critique.

One might object to this line of reasoning: why can we not suppose that *C* *has* been replaced by *C'*, but is temporarily restored for the performance of the argument by analogy?<sup>22</sup> The response is that we *can* suppose this. But, if performing an argument by analogy requires temporarily restoring the explicandum, then it would be misleading to suggest that the argument by analogy is making use of a *full* explication. Even if theorists use *C'* in place of *C* in general, the two concepts are being used

---

<sup>22</sup> This objection was raised by an anonymous referee.

side-by-side in the specific context in which the argument by analogy is taking place—a context which involves applications of both concepts to the relevant subject matter. Arguments by analogy make use of *partial* explications, even if the explicanda have been replaced—and thus have been fully explicated—for *other* theoretical purposes.

Suppose now that we have Strawson’s critique of Oppenheimian explication in mind. First, according to Strawson, where Oppenheimian explication is used to help solve a philosophical problem, what ‘gives life and meaning’ to that solution will be the extra-systematic remarks that explain the respects in which the explicatum is similar to the explicandum. Nothing has been said by proponents of arguments by analogy to undermine this claim. For example, it remains open for Strawson to claim that the plausibility of Maher’s arguments (1)–(3) rely pivotally on Maher’s ability to spell out precisely the respects in which CONFIRMATION’ is similar to CONFIRMATION. Indeed, without such details, it would be unclear why a translation of Maher’s proof of the falsity of *N*’ into ‘qualitative explicandum terms’ would yield a good direct argument against *N*.

Nonetheless, arguments by analogy do give us reason to think that Strawson has exaggerated his case. Arguments by analogy show that Oppenheimian *partial* explication can play *more* than an ancillary role in solving philosophical problems. Partial explications (whether Carnapian or Oppenheimian) play an *essential* role in arguments by analogy: the introduction of a precise explicatum *C*’ must precede the derivation of *F*(*C*’), which is an essential component of the argument by analogy. Thus, *pace* Strawson, partial explication can play an essential, more-than-ancillary role in solving philosophical problems.

### 4.3 Close Similarity

It has been suggested, in particular by Jonah Schupbach (2017), that Strawson’s critique principally concerns the similarity between explicatum and explicandum.<sup>23</sup> According to Schupbach, “Strawson maintains that a logically or scientifically precise explicatum can never illuminate an imprecise

---

<sup>23</sup> Strawson’s critique is also explicitly tied to similarity by Brun (2016: 2018–2019) and Dutilh Novaes and Reck (2017: 202–203).

explicandum because they are too dissimilar” (p. 684). Here, Schupbach takes an explicandum and explicatum to be similar to “the extent to which properties of the explicandum also hold true of the explicatum” (p. 688)—where the properties of an explicandum are manifested in the ordinary judgements of those who use that concept, and the properties of an explicatum can be derived through analysis of its definition or rules of use.<sup>24</sup> To respond to the critique so-construed, one must demonstrate how to “establish the similarity of explicatum to explicandum that is necessary in cases where one intends for the former to illuminate the latter” (p. 685).

Schupbach argues that experimental methods can play a central role in demonstrating that, in relevant respects, the explicatum bears close similarity to explicandum.

[E]mpirical research provides us with crucial information for assessing more directly just how well a particular explication does with regards to explication’s similarity desideratum. By observing and surveying how particular groups of people apply the relevant term(s) in the relevant context(s), we gain [...] evidence of how closely overall (in the contexts tested) any particular formal explicatum corresponds to the explicandum. (p. 689)

For example, Schupbach provides initial experimental evidence (first described in his 2011) that the quantitative concept  $\mathcal{E}(e, h)$ —the explanatory strength of  $h$  over evidence  $e$ —bears close similarity to the ordinary concept EXPLANATORY POWER. Here:

$$\mathcal{E}(e, h) = \frac{Pr(h|e) - Pr(h|\neg e)}{Pr(h|e) + Pr(h|\neg e)}.$$

Participants were told the initial contents of two urns, A and B, which contained different numbers of black and white balls, and then shown balls being drawn without replacement from a randomly selected urn (participants did not know which). After each draw, the participants were asked to judge the power of two potential explanations—‘balls are being drawn from A’ and ‘balls are being drawn from B’—of the results of the draws thus far. These judgements were then compared to the

---

<sup>24</sup> Although Schupbach is not explicit about this, I understand the idea to be that the ‘properties’ in question include, in particular, extension in various (everyday) actual and hypothetical scenarios, and inferential connections with other related concepts. If this is right, then Schupbach’s understanding of simplicity differs from Carnap’s (see footnote 6).

corresponding value of  $\mathcal{E}(e,h)$ , both as determined by objective probability and by the participants' subjective probabilities. Schupbach summarises the results thus:

results derived from  $\mathcal{E}$  sit closer, on average, to participant judgments of explanatory power, and the mean residual corresponding to  $\mathcal{E}$  is closer to the ideal value of zero than that corresponding to any other measure.<sup>25</sup> (2017: 694)

(See Schupbach 2011 and 2017: 692–695 for details.) Schupbach takes this to be clear—albeit not conclusive—evidence that the explicatum bears close similarity to the explicandum. Given sufficient evidence of this kind, close similarity could be established.

Schupbach has spelt out a powerful experimental method for establishing that, in one good sense, an explicatum is similar to an explicandum in relevant respects. However, given that (as seen in §3) Strawson's critique does not rely on the claim that we cannot determine whether explicata are similar to explicanda, Schupbach's method does not automatically constitute a response to Strawson's critique. Let's examine the issue more closely.

Firstly, if one is using Carnapian explication, then one will prioritise fruitfulness for science over similarity. As such, the kind of empirical work advocated by Schupbach will be trumped by the fruitfulness requirement.<sup>26</sup> Strawson can thus maintain that the explicatum in a Carnapian explication will not give rise to the relevant conceptual imbalance—and thus that Carnapian explication is not helpful in solving philosophical problems.

The situation is less clear in the case of Oppenheimian explication. One line Strawson could take is that Schupbach's experimental method in fact *supports* his claim that the extra-systematic remarks are what 'give life and meaning' to an attempt to solve a philosophical problem. Let me

---

<sup>25</sup> The *mean residual corresponding to a measure of explanatory power* is the average difference between the value of that measure and the participants' judgements of explanatory power. The participants recorded their judgements by placing a mark along a line, with one end of the line representing an extremely poor explanation and the other end representing an extremely good explanation; these marks were later converted to numerical values (between –1 and 1) for analysis. See Schupbach 2011 for details.

<sup>26</sup> For this reason, Schupbach only sees his approach as being relevant to Oppenheimian explication.

explain. Schupbach intends his empirical work to demonstrate the respects in which the explicata are similar to the explicanda. As such, for Strawson, that empirical work is *not* a constituent of the explication *per se*, but is rather a part of the extra-systematic remarks. But then, plausibly, Schupbach's experimental method merely serves to emphasise how much clarificatory work can in fact be done by a full set of extra-systematic remarks. If Schupbach were to use  $\mathcal{E}(e,h)$  to solve a philosophical problem arising from consideration of EXPLANATORY POWER, then—modulo the comments made in the previous section—it would be open for Strawson to maintain that the principal work is done by Schupbach's empirical work, and thus the extra-systematic remarks, rather than the explication *per se*.

What we say here, however, ultimately depends on what counts as a constituent of an explication. Strawson assumes that explanations of how explicata are similar to explicanda, i.e. his extra-systematic remarks, are *not* constituents of corresponding explications. This assumption is natural if one conceptualises explication as an *action* of replacement.<sup>27</sup> If explication is conceptualised in this way, then a theorist *performs an explication* by replacing an old concept with an appropriate counterpart, and an explication can be *evaluated* according to the how well the explicatum satisfies the four requirements of similarity, precision, fruitfulness and simplicity. Discussion of those requirements will not be a part of the explication *per se*, but rather a *justification* for it; and so, *a fortiori*, any extra-systematic remarks will not be a constituent of the corresponding explication. From this perspective, as noted above, Schupbach's experimental method plausibly *supports* Strawson's critique of Oppenheimian explication.

However, one need not conceptualise explication in this way. Instead, one might conceptualise explication as a *process* that effects a replacement.<sup>28</sup> So conceptualised, explication may have multiple steps, perhaps: first, clarify the sense in which explicanda are to be explicated; second, construct explicata that satisfy the four requirements of similarity, precision, fruitfulness and simplicity to a sufficient degree; and, third, replace the explicanda with the explicata. It is natural,

---

<sup>27</sup> For example, this seems to be how Brun conceptualises the method. See his 2016: 1220.

<sup>28</sup> For example, see Dutilh Novaes and Reck 2017: 199.



here, to see discussion of the four requirements *not* as justificatory, but rather as *part of the process*: in particular, consideration of the requirements is an essential part of completing the second step. As such, any extra-systematic remarks *will* be a constituent of the explication *per se*.

How, if explication is a process, does Schupbach's experimental method bear on Strawson's critique of Oppenheimian explication? Not at all. The reason is that, from this perspective, Strawson's critique of Oppenheimian explication seems somewhat confused from the outset: it is nonsensical to downplay the importance of a process by emphasising the importance of *part* of that process, regardless of whether experimental methods can be used at that part of the process. So, if explication is conceptualised as a process, then Strawson's critique of Oppenheimian explication fails—but independently of Schupbach's response.

I do not at this stage, however, want to make too much of the distinction between explication-as-action and explication-as-process. Strawson's critique targets the acts of introducing an explicatum and replacing the explicandum, regardless of whether such acts are to be identified as (partial and full) explications. Where exactly we draw the boundaries of an explication seems little more than theoretical bookkeeping. For the sake of argument, then, I henceforth assume that explication is an action; that is, that the extra-systematic remarks are *not* part of the explication.

Thus, while Schupbach has provided us with a method for demonstrating that, in one good sense, explicata bear close similarity to explicanda, this does not constitute a response to Strawson's critique. First, Schupbach's experimental method is largely irrelevant to Strawson's objection to the use of Carnapian explication in solving philosophical problems. Second, on at least one way of conceptualising explication, Schupbach's experimental method is plausibly construed as adding support to Strawson's critique of Oppenheimian explication.

## **5. The Role of Explication in Philosophical Methodology**

Strawson's critique is a challenge to those who would afford explication a role in philosophical methodology. We have already seen that the critique does *not* seek to show that explication is straightforwardly irrelevant; it is in fact compatible with explication playing such a role. First, from

the outset, Strawson allows that Oppenheimian explication can help to solve philosophical problems, although he downplays the contribution of explications *per se*. And, second, Strawson builds the critique on a specific account of philosophical problems, and the critique may not fully generalise. Moreover, *pace* Strawson, we have seen a style of argument, the argument by analogy, in which *partial* explication appears to play an essential role.

In this section, I offer a more systematic response to Strawson's critique. I briefly refine the account of explication on the table in §5.1, before tackling the critique head-on in §5.2 and generalising to other conceptions of philosophical problems in §5.3. All of this will help to clarify the role that explication can play in philosophical methodology.

### **5.1 Fruitfulness versus the Carnapian/Oppenheimian Distinction**

Recall that Carnapian explication involves designing explicata to be fruitful for the (empirical or logicomathematical) sciences; and Oppenheimian explication involves designing explicata to be maximally similar to explicanda in relevant respects. The distinction is useful for Strawson's (and Schupbach's) specific purposes. But, for two reasons, it is not so helpful when thinking about explication more generally.

First, very briefly, the distinction is not exhaustive. For example, one might design an explicatum to be fruitful for philosophy, or to have an optimal balance of fruitfulness and similarity.

Second, more fundamentally, the apparent need for the distinction rests on an unhelpfully narrow conception of fruitfulness. Both Strawson (1963: 504) and Schupbach (2017: 675) think of fruitfulness in broadly Carnapian terms: an explicatum is fruitful insofar as it facilitates the formulation of empirical laws or logical theorems. With fruitfulness so construed, a theorist who seeks to illuminate the explicandum must prioritise similarity over fruitfulness.

However, Carnap's account of fruitfulness is inappropriate for many areas of theoretical inquiry. For one example, the astronomers' explication of PLANET appears to have been fruitful *not* because it features in laws or theorems, but because it facilitated the development of a principled,

standardised taxonomy of celestial objects.<sup>29</sup> Others have made a similar observation, for example Kitcher argues that Carnap’s account of fruitfulness is “deeply problematic for the biological, earth and human sciences” (2008: 115), and Shepherd and Justus argue that it is inappropriate for normative concepts (2015: 392f). To avoid such concerns, I suggest that we tie fruitfulness to the specific theoretical aims that led to the introduction of the explicatum: if a theorist seeks to answer various theoretical questions, then the explicatum will be fruitful insofar as it allows one to answer such questions (Kitcher 2008); and if a theorist seeks new knowledge about some phenomenon, then the explicatum will be fruitful insofar as it leads to such knowledge (Dutilh Novaes and Reck 2017: 206); and so on.

From this perspective, there is no need to draw a distinction between Carnapian and Oppenheimian explication. If a theorist aims her explicatum to be useful for science, then the explicatum will be fruitful insofar as it is useful for science. Alternatively, if a theorist aims to illuminate an explicandum, then the explicatum will be fruitful insofar as it facilitates such illumination; if illumination requires the explicatum to bear close similarity to the explicandum, then this will be subsumed under the *fruitfulness* requirement. Likewise for other theoretical aims: whatever the aim, it is *fruitfulness* that is to be prioritised. I recommend, then, that we drop the distinction between Carnapian and Oppenheimian explication.

## 5.2 Assessing Strawson’s Critique

In this subsection, I tackle Strawson’s critique head on, keeping in mind his underlying account of philosophical problems, introduced in §3.1. I will use the sceptical imbalance as a running example.

There are three variables that affect the force of the critique as applied to the sceptical imbalance. First, the explicandum might be KNOWLEDGE, or it might be something else, say, SENSORY EXPERIENCE. Presumably, if the explicandum is the latter, the intention is that the explicatum clarify, or help draw out, the sceptical imbalance. Accordingly, I assume that explicating SENSORY EXPERIENCE does *not* affect whether or not the sceptical imbalance arises. Second, in designing the

---

<sup>29</sup> See Pinder 2017a.

explicatum, the explicator might aim to build a concept that is any number of things, such as *useful for contemporary psychology*, *useful for contemporary epistemology*, *maximally similar to the explicandum*, or something else. And, third, the explication may be full or partial.

First, then, suppose that a philosopher explicates SENSORY EXPERIENCE. Then, *ex hypothesi*, the sceptical imbalance still arises—regardless of whether the philosopher designed the explicatum for this specific purpose or whether she (say) adopted an explicatum originally designed for use in psychology or epistemology, and regardless of whether the explicandum is replaced. As such, the explication will be immune to Strawson’s charge of changing the subject. However, such an explication plausibly does play only an ancillary role in any subsequent solution to the philosophical problem: the explication helps us to understand the problem, rather than to solve it.

Suppose instead that a philosopher explicates KNOWLEDGE, and is to replace the explicandum. That is, the philosopher will *fully* explicate KNOWLEDGE. Then, for Strawson, a key question is whether the explicatum also gives rise to the sceptical imbalance. If it *does*, then Strawson’s grounds for claiming a change of subject will be undercut. In such a case, the explicit precision afforded by the explicatum could potentially make it possible to resolve the sceptical imbalance. If this were so, then the explication would certainly not be ancillary, instead playing an essential role in solving the philosophical problem.

For example, suppose that one explicates KNOWLEDGE, with the explicatum defined by an appropriate contextualist account of knowledge.<sup>30</sup> Call the explicatum “C-KNOWLEDGE”. This explicatum plausibly *would* give rise to the sceptical imbalance, in something like the following sense: if we reconstruct the derivation of the sceptical imbalance by replacing each use of KNOWLEDGE with an appropriate use of C-KNOWLEDGE, then each step in that derivation will be valid in a (salient) context of utterance. One might then hope to resolve the sceptical imbalance by pointing to the fact that the reconstructed derivation *as a whole* is invalid from any given context of utterance:

---

<sup>30</sup> See e.g. DeRose 1995; Lewis 1996.

our mistake, in deriving the sceptical imbalance, is in failing to recognise shifts in contexts of utterance.<sup>31</sup>

It will be a matter of judgement whether an explicatum gives rise to a given conceptual imbalance (and, *a fortiori*, whether I am right to suggest that C-KNOWLEDGE gives rise to the sceptical imbalance). The judgement should turn principally on whether the derivation of the conceptual imbalance can be naturally reconstructed using the explicatum, in such a way that it is clear why each step in the derivation might seem sound and why the conclusion might seem unacceptable. If the derivation can be naturally reconstructed in this way, we may judge that the explicatum does give rise to the conceptual imbalance. This is an issue to be considered on a case-by-case basis.

If the explicatum replacing KNOWLEDGE does *not* give rise to the sceptical imbalance, then the explication will be open to the charge of changing the subject. Strawson takes this to be the inevitable outcome of using an explicatum designed for use in science, as it will have a purpose too narrow to give rise to the sceptical imbalance. For example, we might imagine that a philosopher adopts an explicatum designed for fruitfulness in psychology—perhaps TACIT KNOWLEDGE, defined roughly as a body of information (construed non-factively) stored in a module, such as our native language or the content stored in the visual system. If KNOWLEDGE is replaced by TACIT KNOWLEDGE, then the sceptical imbalance is likely not to arise: in ordinary situations, our visual system plausibly contains tacit knowledge of our immediate surroundings whether or not we can distinguish current experience from a vivid dream. I am not sure whether this point generalises to *all* explicata designed for use in science, but a tempered analogue of Strawson’s objection here appears correct: fully explicating a concept that gives rise to a philosophical problem, where the explicatum has been designed for use in science, will *typically* constitute a change of subject.

One might be tempted to respond: “*No philosopher would seek to solve a philosophical problem by replacing the concepts of interest with concepts designed for use in science in this way! So, as an objection to explication, this is spurious!*” However, firstly, there are readings of, for

---

<sup>31</sup> This is *not* to say that contextualism about knowledge solves the problem of scepticism as standardly construed. See e.g. Kornblith 2000.

example, Carnap (1963: 934) and Quine (1969; 1975; cf. Fumerton 1994) on which they are doing just that. So the objection is not spurious. Secondly, even for more sophisticated uses of explication, there is something to be learnt. The objection arises whenever the explicatum does *not* give rise to the conceptual imbalance in question. This is likely to happen in the present example if the explicatum for KNOWLEDGE is designed for use in psychology; but it *might* also happen if the explicatum is designed for use in epistemology, or perhaps even if it is designed to be maximally similar to our ordinary concept. In such cases, replacing KNOWLEDGE with the explicatum will not serve to resolve the sceptical imbalance.

For example, a philosopher might replace KNOWLEDGE with the explicatum RELIABILIST KNOWLEDGE. Let us assume that the explicatum does not give rise to the sceptical imbalance, as (say) my belief that I am sitting in my armchair may have been formed by a reliable belief-forming-mechanism, and thus may fall under RELIABILIST KNOWLEDGE, even if my experience is indistinguishable from a vivid dream. Such a replacement *per se*, however, does not constitute a solution to the philosophical problem at hand: it does not serve to explain *why*, when reflecting on the concepts involved in the original formulation of the sceptical imbalance, it seems that I do *not* know that I am sitting in my armchair.

Suppose now that the philosopher does not *fully* explicate KNOWLEDGE, but instead *partially* explicates it. If the explicatum does give rise to the sceptical imbalance, then the points above about C-KNOWLEDGE will straightforwardly carry over. So let us assume that the explicatum does *not* give rise to the sceptical imbalance. Two points immediately follow: firstly, despite not giving rise to the sceptical imbalance, the explication will not change the subject as the explicandum may still be used in the relevant theoretical contexts; and, secondly, the explication will open up the opportunity for arguments by analogy, and thus may play more than an ancillary role in solving the philosophical problem.

For example, a philosopher might *partially* explicate KNOWLEDGE with the explicatum RELIABILIST KNOWLEDGE. Then (rather than simply eliminating the explicandum), she might point to appropriate similarities between KNOWLEDGE and RELIABILIST KNOWLEDGE in order to resolve the sceptical imbalance. To do this, she might argue by analogy (say) that my belief that I am sitting in

my armchair falls under not only RELIABILIST KNOWLEDGE, but also KNOWLEDGE.<sup>32</sup> If successful, then the explication would have played an essential role in solving the philosophical problem.

In such a case, a careful, detailed explanation of the relation between explicatum and explicandum is pivotal. We need to understand how RELIABILIST KNOWLEDGE and our everyday concept KNOWLEDGE are related to each other, and how they are both related to the sceptical imbalance: we need an explanation of why and how we can learn about our everyday concept from the introduced concept. One might hope to make use of Schupbach's experimental approach (discussed in §4.3), or something else. Either way, Strawson is *right* to emphasise the importance of the extra-systematic remarks—even if he is *wrong* to downplay the importance of explication.

### 5.3 Alternative Conceptions of Philosophical Problems

So far, we have followed Strawson's account of philosophical problems as conceptual imbalances. For Strawson, philosophers seek to understand how our concepts give rise to apparent paradoxes and perplexities, and how we can make sense of such paradoxes and perplexities. This may be a good model for some philosophical problems, but it is inadequate as a general account: I see little reason to insist that *all* philosophical problems be characterised as paradoxes and perplexities. Moreover, as suggested above, Strawson's critique is unlikely to fully generalise. It is not possible here to develop and examine a full taxonomy of types of philosophical problem. Instead, let me consider the role that explication might play with respect to two broad, alternative ways that one might think about the problem of scepticism. This will serve to illustrate some different ways in which the above discussion may generalise.

First, one might take a philosophical problem *not* to concern some given concept, but rather to concern whatever in the world that concept is about. For example, one might object to the idea that the problem of scepticism demands merely a clarification of KNOWLEDGE, claiming instead that it is a problem about knowledge.<sup>33</sup> This approach to philosophical problems is most naturally tied to

---

<sup>32</sup> Something like this seems to be the idea in Olsson 2015, although Olsson does not explicitly discuss scepticism.

<sup>33</sup> See e.g. Williamson 2007.

semantic externalism, which we temporarily put aside in §3.1. Along these lines, one might reason thus: there is something we call “knowledge”; we value that something and would like to understand what it is; the sceptical imbalance highlights a flaw in our understanding of it; and so the sceptical imbalance is a problem to solve in pursuit of a full theory of what we call “knowledge”. On this approach, our everyday concept KNOWLEDGE has (via some appropriate reference-fixing mechanism) latched onto something in the world (such as a kind of mental state), and it is the latter, not the former, which is of philosophical interest.

With the problem so construed, a number of points can be made about the use of explication. Firstly, there is no in-principle objection to the use of explication to refine related concepts such as SENSORY EXPERIENCE. Such an explication will be desirable if performed with a relevant theoretical aim, such as to bring knowledge into better focus, but is likely to remain ancillary. Secondly, there is no in-principle objection to a *partial* explication of KNOWLEDGE: so long as we can still use the explicandum to pick out knowledge, there will be no change of subject. There is thus the possibility that arguments by analogy play an important role in making discoveries about knowledge and, *a fortiori*, that explication play an essential, not merely ancillary, role. Thirdly, a *full* explication of KNOWLEDGE is liable to be *accused* of changing the subject unless good evidence can be given that reference has been preserved across the explication. If such evidence can be provided, then the explication might serve, for example, to eliminate some of the confusions inherent in our ordinary way of thinking about knowledge. However, the provision of such evidence may well depend, at least to some extent, on an independent grasp of knowledge—in which case the explication *per se* (or, at least, the replacement) would be ancillary. So a variant of Strawson’s critique holds for a *full* explication of KNOWLEDGE: the explication will be open to the charge of changing the subject, or else be ancillary.

Another way to think of a philosophical problem is as a problem targeting some general *phenomenon*. From this perspective, the relevant philosophical problem does not target a specific concept, nor the extension of that concept; rather, it targets some more general phenomenon to which the concept and its extension pertain in some way. Thus, rather than thinking of the problem of scepticism as targeting KNOWLEDGE or knowledge, we may think of it as targeting the general



phenomenon whereby people can be in various epistemic states. For example, in response to the problem of scepticism, one might reason thus: people can be in a variety of epistemic states in relation to themselves, each other and the outside world; to make sense of this, we need a principled taxonomy of epistemic states; as a first approximation, we may distinguish belief, justified belief, knowledge, etc.; but when we look closer at actual epistemic behaviour—how epistemic states are initially formed, how they are revised in light of new information, what actions they lead to, whether these practices are conducive to the satisfaction of desires, world-to-mind states, pro-attitudes or something else, etc.—it may turn out that we need to improve upon our initial taxonomy; the sceptical imbalance highlights a general difficulty in delineating any factive state, whether we call it “knowledge” or something else; and so the sceptical imbalance is a problem that may bear on our attempt to find a principled way to distinguish different kinds of epistemic state. On this approach, the problem of scepticism is principally a technical challenge in the development of a principled taxonomy of epistemic concepts.

Again, a number of points can be made about the use of explication in solving the philosophical problem so construed. Firstly, in order to refine and clarify our initial taxonomy, it will almost certainly be helpful to *partially* explicate SENSORY EXPERIENCE, KNOWLEDGE, and a variety of other epistemic concepts. This will allow the construction of competing systems of concepts, whose theoretical and explanatory efficacy can be tested against the phenomena. Secondly, it cannot be decided in advance whether those concepts will, or ought to, be *fully* explicated. Some of the constructed systems of epistemic concepts may include explicata of SENSORY EXPERIENCE and KNOWLEDGE, but other systems may not.<sup>34</sup> One way or another, one such system of concepts may emerge as dominant, replacing earlier taxonomies. If that system contains an explicatum of SENSORY EXPERIENCE or KNOWLEDGE, then that concept will have been fully explicated; otherwise not.

Thirdly, with the philosophical problem so construed, a charge of changing the subject will not apply in the vast majority of cases, whether or not the explicata give rise to the sceptical imbalance. So far as I can see, the charge would only be appropriate if a theorist sought to fully

---

<sup>34</sup> See, e.g., Churchland 1981 and the discussion thereof in Pinder 2017b.

explicate (say) KNOWLEDGE in such a way that the explicatum was no longer plausibly an epistemic concept—say, for example, if one replaced KNOWLEDGE with the concept TRUTH—but it is difficult to see a serious epistemologist performing such an explication. Fourthly, explication *per se* would not play a mere ancillary role in this case; rather, it would play a driving role in the construction of a taxonomy appropriate for epistemic theorising. Perhaps some particular (partial) explications would only play an ancillary role—for example, those that had little effect on the development of the taxonomy that ultimately becomes dominant—but this is not a criticism of the role of explication *per se*.

To reiterate, this is not intended as an exhaustive taxonomy of types of philosophical problem. For example, one might think that ethical questions do not fit into any of the three types of philosophical problem I have highlighted: questions about how we ought to act may not be merely conceptual, nor concern the extension of ‘ought’, nor, in any helpful sense, be about ‘ethical phenomena’. I am not sure that such a thought is right but, even if it is, it is no objection to the above discussion. The underlying theme in this section is that explication *can* play an important role in philosophical methodology *in the right circumstances*. To determine whether it will be helpful in a given case, one must look closely at the problem at hand and at what exactly the explication is supposed to achieve. I have done this, abstractly, for three different ways of conceiving of philosophical problems. This explicitly delimits the role that explication can play in solving philosophical problems conceived along those lines and, additionally, provides a model for how to delimit the role that explication can play in philosophical problems more generally.

Finally, note that I am *not* suggesting that each philosopher must adopt just one conception of philosophical problems, nor that there is a right conception for a given problem. I am happy, for example, for a single epistemologist to undertake three separate projects: to resolve the sceptical imbalance; to fully understand the nature of what we ordinarily call “knowledge”; and to develop a system of concepts that is specifically designed to capture and explain epistemic phenomena. In each case, she may decide to explicate KNOWLEDGE for the respective purpose. So long as, in each case,

her use of explication stays within the limits set above—something which is likely to require her to explicate KNOWLEDGE in three different ways—I have no objection.<sup>35</sup>

## 6. Concluding Remarks

Strawson did not object to the general use of explication in philosophy, and certainly not on the grounds that it must change the subject. Nor, as others have pointed out, is such a general objection sound. Strawson *does* claim, however, that, in the service of resolving a conceptual imbalance, one would change the subject by replacing a concept essential to its formulation with a concept designed to be fruitful for science. A tempered variant of that objection is sound, but limited. Ultimately, a more systematic investigation is required to delimit the viable use of explication in philosophical methodology. Whether or not explication changes the subject depends on how the explication is performed, how it is to be used, and how the philosophical problem being tackled has been construed.

Strawson also objects that, if one seeks to maximise the similarity between explicatum and explicandum, then explication will play only an ancillary role in solving philosophical problems. Extant responses in the literature—especially Maher’s arguments by analogy—have already shown this objection to be flawed. Again, a more systematic investigation is required to establish when and how explication can play an *important* role in philosophical methodology. This will likewise depend upon how the explication is performed, what it is to be used for, and how the philosophical problem being tackled has been construed.

In responding to Strawson’s critique, I have made some brief but systematic remarks about the use of explication in philosophical methodology. One thing that the discussion makes clear is that the viability and value of explication will be sensitive to the details of any given case. Before using explication in philosophy, then, it is essential to be explicit about exactly how one is using it, for what purpose, and how this is relevant to the philosophical problem at hand. Explication can be a powerful

---

<sup>35</sup> See Kitcher 2008 for a discussion of how explication may lead to pluralities of concepts in broadly this way.

tool in philosophical methodology but, in any particular case, its applicability ought not to be assumed.<sup>36</sup>

## 7. References

- Brun, G. (2016). Explication as a method of conceptual re-engineering. *Erkenntnis*, 81(6), 1211–1241.
- Carnap, R. (1947). *Meaning and Necessity*. Chicago: University of Chicago Press.
- . (1950). *Logical Foundations of Probability*. Chicago: The University of Chicago Press.
- . (1963). Replies and systematic expositions. In Schilpp 1963: 859–1013.
- Churchland, P. (1981). Eliminative materialism and the propositional attitudes. *Journal of Philosophy*, 78(2), 67–90.
- DeRose, K. (1995). Solving the Skeptical Problem. *Philosophical Review*, 104(1), 1–52.
- Dutilh Novaes, C., & Reck, E. (2017). Carnapian explication, formalisms as cognitive tools, and the paradox of adequate formalization. *Synthese*, 194(1), 195–215.
- Fumerton, R. (1994). Skepticism and Naturalistic Epistemology. *Midwest Studies in Philosophy*, 19, 321–340.
- International Astronomical Union. (2006). Resolution B5: Definition of a Planet in the Solar System. URL: [https://www.iau.org/static/resolutions/Resolution\\_GA26-5-6.pdf](https://www.iau.org/static/resolutions/Resolution_GA26-5-6.pdf). Retrieved 4th April 2017.
- Justus, J. (2012). Carnap on concept determination: methodology for philosophy of science. *European Journal for Philosophy of Science*, 2(2), 161–179.

---

<sup>36</sup> I thank two anonymous referees for this journal, one of whom in particular provided very helpful and insightful comments. I thank the organisers and participants of the *Philosophical Methods* workshop held in Essen in June 2016. As a result of the excellent papers and discussion at that workshop, my approach to Strawson's critique has evolved significantly from the paper presented there. I particularly benefitted from discussion with Magdalena Balcerak Jackson, Georg Brun, Catarina Dutilh Novaes, Anna-Maria Asunta Eder, Hannes Leitgeb and Timothy Williamson. I also thank Nick K. Jones and Jonah Schupbach, whose comments on other projects have significantly influenced this work.

- Kemeny, J.G., & Oppenheim, P. (1952). Degree of factual support. *Philosophy of Science*, 19(4), 307–324.
- Kitcher, P. (2008). Carnap and the caterpillar. *Philosophical Topics*, 36(1), 111–127.
- Kornblith, H. (2000). The Contextualist Evasion of Epistemology. *Philosophical Issues*, 10, 24–32.
- Kripke, S. (1980). *Naming and Necessity*. Cambridge, MA: Harvard University Press.
- Lewis, D. (1996). Elusive Knowledge. *Australasian Journal of Philosophy*, 74(4), 549–567.
- Loomis, E., & Juhl, C. (2006). Explication. In S. Sarkar & J. Pfeifer (Eds.) *The Philosophy of Science: An Encyclopedia (Volume 1)* (pp. 287–294). New York: Routledge.
- Maher, P. (2004). Probability captures the logic of scientific confirmation. In C. Hitchcock (Ed.) *Contemporary Debates in Philosophy of Science* (pp. 69–93). Oxford: Blackwell.
- . (2007). Explication defended. *Studia Logica*, 86(2), 331–341.
- Olsson, E. (2015). Gettier and the method of explication: a 60 year old solution to a 50 year old problem. *Philosophical Studies*, 172(1), 57–72.
- Pinder, M. (2017a). Does experimental philosophy have a role to play in Carnapian explication? *Ratio*, doi: 10.1111/rati.12164.
- . (2017b). The explication defence of arguments from reference. *Erkenntnis*, doi: 10.1007/s10670-016-9868-9.
- Putnam, H. (1975). The meaning of ‘meaning’. *Minnesota Studies in the Philosophy of Science*, 7, 131–193.
- Quine, W.V.O. (1969). Epistemology Naturalized. In his *Ontological Relativity and Other Essays* (pp. 69–90). New York: Columbia University Press.
- . (1975). The Nature of Natural Knowledge. In S. Guttenplan (Ed.) *Mind and Language* (pp. 67–81). Oxford: Clarendon Press.
- Schilpp, P. (Ed.) (1963). *The Philosophy of Rudolf Carnap*. LaSalle, IL: Open Court.
- Schupbach, J. (2011). Comparing probabilistic measures of explanatory power. *Philosophy of Science*, 78(5), 813–829.
- . (2017). Experimental explication. *Philosophy and Phenomenological Research*, 94(3), 672–710.

- Shepherd, J., & Justus, J. (2015). X-phi and Carnapian explication. *Erkenntnis*, 80(2), 381–402.
- Strawson, P. F. (1963). Carnap's views on constructed systems versus natural languages in analytic philosophy. In Schilpp 1963: 503–518.
- Williamson, T. (2007). *The Philosophy of Philosophy*. Oxford: Blackwell.

## 8. Appendix

Summary of results of discussion in §5

Explicandum	Explicatum	Type	Philosophical Problem		
			Resolve Sceptical Imbalance	Understand <i>Knowledge</i>	Taxonomise Epistemic Phenomena
KNOWLEDGE	C- KNOWLEDGE	Partial	Possibly essential	Possibly essential	Possibly essential
		Full	Possibly essential	Ancillary or open to charge of changing subject	Possibly essential
	RELIABILIST KNOWLEDGE	Partial	Possibly essential	Possibly essential	Possibly essential
		Full	Change of subject	Ancillary or open to charge of changing subject	Possibly essential
	TACIT KNOWLEDGE	Partial	Possibly essential	Possibly essential	Possibly essential
		Full	Change of subject	Ancillary or open to charge of changing subject	Possibly essential
	TRUTH	Partial	Possibly essential (?)	Possibly essential (?)	Possibly essential (?)
		Full	Change of subject	Ancillary or open to charge of changing subject	Change of subject
SENSORY EXPERIENCE	SENSORY EXPERIENCE/ EXPERIENCE'	Partial	Ancillary	Ancillary	Possibly essential
		Full	Ancillary	Ancillary	Possibly essential

Figure 1. Summary of the role that explication can play in philosophical methodology. *C-KNOWLEDGE* represents an explicatum on which the sceptical imbalance arises, *RELIABILIST KNOWLEDGE* an explicatum (designed for use in epistemology) on which the sceptical imbalance does not arise, *TACIT KNOWLEDGE* an explicatum designed for use in science, and *TRUTH* an explicatum that is not an epistemic concept. The “(?)” indicates that, as *TRUTH* is not an epistemic concept, there are grounds independent of Strawson’s critique for doubting the efficacy of partially explicating *KNOWLEDGE* with *TRUTH* in order to solve a philosophical problem.